Available online at www.sciencedirect.com

ScienceDirect

Journal of Taibah University for Science 10 (2016) 787–796

Journal of Taibah University for Science

www.elsevier.com/locate/jtusci

Combining DFT and QSAR studies for predicting psychotomimetic activity of substituted phenethylamines using statistical methods

A. Aouidate [a,*], A. Ghaleb [a], M. Ghamali [a], S. Chtita [a], M. Choukrad [a], A. Sbai [a], M. Bouachrine [b], T. Lakhlifi [a]

[a] *Molecular Chemistry and Natural Substances Laboratory, Faculty of Science, Moulay Ismail University, Meknes, Morocco*
[b] *ESTM, Moulay Ismail University, Meknes, Morocco*

Received 25 April 2016; received in revised form 29 June 2016; accepted 1 July 2016
Available online 20 July 2016

Abstract

The DFT-B3LYP method, with the base set 6-31G (d) was used to calculate electronic and charge descriptors. The present study was performed using principal component analysis (PCA), multiple linear regression analysis (MLR) and non-linear multiple regression analysis (MNLR) to predict unambiguous QSAR models of 46 substituted phenethylamines toward psychotomimetic activity. Results showed that the MLR and MNLR predict activity in a satisfactory manner. But among those models, we concluded that the latter one provides a better agreement between calculated and observed values of psychotomimetic activity. Also it shows very good stability towards data variations for the validation methods.

Keywords: DFT method; Phenethylamines; Psychotomimetic activity; QSAR model

Contents

[*] Corresponding author.
E-mail address: a.aouidate@hotmail.fr (A. Aouidate).

Peer review under responsibility of Taibah University

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jtusci.2016.07.001

1. Introduction

Ideal psychotomimetics are those agents which produce changes in thought, mood, and perception, with little memory impairment, little stupor, narcissism or excessive stimulation, minimal autonomic side effects, and are not addicting, as they are defined by Hollister [1]. Their psychotomimetic activity is expressed in relation to the chosen standard mescaline (UM) and defined as the ratio of the effective dose of mescaline to the effective dose of the tested compound [2]. Among the well-known psychotomimetics are the phenethylamines [3,4], which will be the subject of the present study.

The phenethylamine derivatives such as, amphetamine, methamphetamine, and mescaline that are known to display psychotomimetic activity, have been studied with different approaches so far [5,6]. As it is difficult to test this type of activity on humans being, theoretical research can circumvent these difficulties and allow obtaining precise data while taking advantage of the rapid progress of computing chemical quantum descriptors, which can be obtained easily from publicly available software. Those can be used to build a quantitative structure activity relationship (QSAR) model to enable calculation of the activity and prediction of the efficacy of new phenethylamines.

The QSAR of phenethylamine psychotomimetics still receives considerable attention because these agents represent a large family of abused substances and continue to be a source of new illicit drugs as witnessed over recent decades [7]. Although QSAR of phenethylamines have previously been developed with steric and lipophilic descriptors [1,8] it is important to extend these with all available data.

The QSAR models described in the previosuly study [8] focus on simple physico-chemical descriptors but are not sufficient for generating comprehensive structure–activity relationships. Electronic descriptors, which can be obtained by calculation, can describe defined molecular activities, and are not restricted to closely related compounds. Therefore, the development of QSAR models in which electronic descriptors are used has great potential [9]. In recent years, some comparative QSAR studies have shown that employing the descriptors calculated using the density functional theory (DFT) method instead of the semi-empirical methods as AM1 or PM3, can improve the accuracy of the results and lead to more reliable QSAR models [10]. Arulmozhiraja and Morita [11] have studied relationships between the various DFT-based descriptors (absolute softness, electronegativity, and electrophilicity index) and the toxicity of 33 polychlorinated dibenzofurans (PCDFs), the results showed a moderate to satisfactory success for the DFT-based reactivity descriptors in the toxicological QSARs. Pasha et al. [12] investigated quantum chemical reactivity descriptors based QSAR models on toxicity of phenol derivatives with AM1, PM3, PM5 and DFT methods, indicating that the DFT method is more reliable than other and has an improved predictive power.

This work is aimed at deriving correlation models, which explain the relationship between the psychotomimetic activity, and the structure of 46 phenethylamines compounds based on electronic, charge and physico-chemical descriptors using several chemometric methods such as principal component analysis PCA, multiple linear regression RML and non-linear regression MNLR.

2. Materials and methods

Psychotomimetic activities of 46 phenethylamines were taken from the literature [2] the activity was expressed as MU (Mescaline Units) and is defined as mole mescaline/mole of the tested phenethylamine. Fig. 1 and Table 1 show the substituted structures of the studied compounds. For modeling, the data set was split into two sets. Thirty five molecules were chosen randomly to represent the quantitative model (Training set) and the rest were used to test the performance of the proposed model (test set). Additionally a leave-one-out

Table 1
Observed activities of studied phenethylamines.

Compound	X	*R*	log *MU*	Compound	X	*R*	log *MU*
1	2,5-OMe,4-Br	Me	2.72	**24**	3,5-OMe,4-OPr	H	0.83
2	2,5-OMe,4-SEt	Me	1.96	**25**	3,4-OMe,5-SEt	H	0.84
3	2,5-OMe,4-Et	Me	2.02	**26**	3-OMe,4-OEt,5-SMe	H	0.84
4	2,5-OMe,4-Pr	Me	1.95	**27**	3,4-OMe,5-SMe	H	0.81
5*	2,5-OMe,4-Me	Me	1.90	**28**	2,3-OMe,4-OCH_2O-5	Me	0.76
6	2,5-OMe,4-S-iPr	Me	1.71	**29**	3-OEt,4-SMe,5-OMe	H	0.66
7*	2,5-OMe,4-Br	H	1.69	**30**	3-OEt,4-SEt,5-OMe	H	0.68
8*	2,5-OMe,4-Bu	Me	1.68	**31**	2,4-OMe	Me	0.67
9	2,5-OMe,4-SMe	Me	1.66	**32**	4-Me	Me	0.59
10	3,5-OMe,4-SEt	H	1.36	**33**	3,5-OMe,4-SBu	H	0.58
11	2,4,5-OMe	Me	1.33	**34**	3,5-OMe,4-$OCH_2C_6H_5$	Me	0.46
12*	2,5-OMe,4-Et	H	1.25	**35***	3-OMe,4-OCH_2O-5	Me	0.43
13*	3,5-OMe,4-SPr	H	1.29	**36**	3-OCH_2O-4	Me	0.41
14	2,5-OMe,4-Me	H	1.27	**37**	3,5-OMe,4-Obu	H	0.38
15	2,5-OMe,3-OCH_2O-4	Me	1.14	**38**	3-SEt,4-OEt,5-OMe	H	0.38
16	2,5-OMe,4-OEt	Me	1.36	**39***	3,4-OEt,5-SMe	H	0.38
17	3,5-OMe,4-SMe	H	1.11	**40**	3,4,5-OMe	Me	0.33
18*	2-OMe,3-OCH_2O-4	Me	1.00	**41**	3,4-OEt,5-OMe	H	0.23
19	3,5-OMe,4-OEt	Me	1.05	**42**	3-OEt,4,5-OMe	H	0.03
20	2-OMe,4-OCH_2O-5	Me	1.00	**43***	3,4,5-OMe	H	0.00
21	2,5-OMe,4-OPr	Me	1.38	**44***	2,3,4-OMe	H	−0.03
22	3,5-OMe,4-OEt	H	0.87	**45***	3,4-OMe	Me	−0.06
23	2,3,4,5-OMe	Me	0.86	**46**	3,4-OMe	H	−0.67

* Test set.

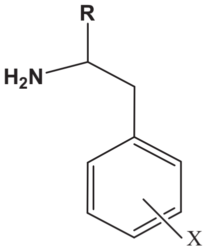

Fig. 1. The chemical structure of the studied compounds.

protocol was performed on the training set for internal validation of the obtained models.

2.1. Molecular descriptors

To describe the compound structural diversity, a total of 13 descriptors encode three important properties have been calculated for each phenethylamine: (a) Charges descriptors, Q_p: the net atomic charge on the para position; Q_{min}: the most negative net atomic charge; (b) electronic descriptors, E_T (eV): the total energy; IP (eV): the ionization potential; E_{HOMO} (eV): the highest occupied molecular orbital energy; E_{LUMO} (eV): the lowest unoccupied molecular orbital energy; DM (Debye): the dipole moment and η (eV): the absolute hardness [13]; were calculated utilizing Gaussian 03 [14] with the DFT calculations.

On the other hand, physico-chemical descriptors were calculated, namely: MW (g/mol): the molecular weight; *D* (g/cm^3): the density; *n*: the refractive index; γ (dyne/cm^3): the surface tension and (log *P*): the octanol/water partition coefficient, utilizing Chemsketch and ChemDraw softwares [15]. Thus, descriptors data matrix of dimension of (46*13) was generated Table 2.

2.2. Methodology

After the calculation of descriptors, a principal component analysis (PCA) [16] was performed to eliminate the correlated descriptors ($R > 0.8$). The remaining descriptors were used to perform an MLR study with backward selection until a valid model including: the critical probability *P*-value <0.05 for all descriptors and for the complete model, The Fisher static, the coefficient of determination, the mean squared error and the multicolinearity test. Later, the chosen variables in the best linear model were exploited to generate the applicability domain, then to evaluate a non-linear model.

2.3. Statistical analysis

In this study XLSTAT version 2013 [17] was used to perform principal component analysis (PCA) multiple

Table 2
The values of Molecular descriptors used in QSAR study.

No.	log *UM*	E_T	IP	E_{HOMO}	η	E_{LUMO}	DM	Q_p	Q_{min}	MW	*n*	γ	*D*	log *P*
1	2.720	−82328.726	7.374	−5.682	5.522	−0.160	3.242	0.046	−0.707	274.154	1.540	37.9	1.324	2.455
2	1.960	−30241.000	7.125	−5.428	5.258	−0.170	1.996	0.135	−0.708	255.376	1.545	42.6	1.080	2.452
3	2.020	−19406.400	6.731	−5.029	5.302	0.273	1.234	0.126	−0.712	223.311	1.509	33.9	0.998	2.778
4	1.950	−20476.100	6.724	−5.037	5.295	0.258	1.283	0.126	−0.709	237.338	1.507	33.9	0.998	3.307
5*	1.900	−18336.700	6.765	−5.009	5.282	0.273	1.117	0.127	−0.710	209.285	1.511	34.2	1.010	2.249
6	1.710	−31310.656	7.238	−4.780	5.066	0.286	2.498	−0.150	−0.712	269.403	1.539	41.2	1.070	2.761
7*	1.690	−86156.656	7.374	−5.687	5.527	−0.160	3.271	0.046	−0.709	260.128	1.548	39.5	1.368	2.146
8*	1.680	−21545.885	6.721	−5.025	5.300	0.275	1.250	0.122	−0.713	251.365	1.504	33.9	0.979	3.386
9	1.660	−29171.207	7.121	−5.453	5.295	−0.158	2.115	−0.131	−0.707	241.350	1.550	43.1	1.100	1.923
10	1.360	−29171.025	6.993	−5.327	5.234	−0.093	4.017	−0.219	−0.713	241.350	1.552	44.2	1.100	1.743
11	1.330	−20382.821	6.830	−5.101	5.576	0.475	3.159	0.349	−0.713	225.284	1.507	34.3	1.048	1.392
12*	1.250	−18336.601	6.802	−5.024	5.301	0.278	1.323	0.125	−0.716	209.285	1.514	34.9	1.012	2.469
13*	1.290	−30240.922	7.156	−5.333	5.236	−0.097	4.084	−0.220	−0.707	255.376	1.547	43.6	1.090	2.272
14	1.270	−17266.813	8.353	−5.142	5.451	0.309	1.618	0.133	−0.714	195.258	1.516	35.3	1.026	1.940
15	1.140	−22396.415	7.075	−5.349	5.755	0.407	1.297	0.273	−0.713	239.268	1.540	43.5	1.184	1.644
16	1.360	−21452.636	6.803	−5.073	5.531	0.458	3.234	0.353	−0.707	239.311	1.504	34.3	1.034	1.921
17	1.110	−28101.400	7.292	−5.423	5.335	−0.088	4.378	−0.223	−0.708	227.323	1.559	45.0	1.120	1.214
18*	1.000	−19280.462	7.265	−5.393	5.876	0.483	1.539	0.326	−0.707	209.242	1.550	45.4	1.175	1.699
19	1.050	−21452.636	7.265	−5.505	5.911	0.406	3.041	0.265	−0.713	239.311	1.504	34.3	1.034	1.571
20	1.000	−19280.543	6.803	−4.968	5.226	0.258	1.505	0.326	−0.715	209.242	1.550	45.4	1.175	1.699
21	1.380	−22522.370	6.911	−5.063	5.528	0.465	3.276	0.352	−0.713	253.337	1.502	34.2	1.022	2.450
22	0.870	−20382.821	7.265	−5.505	5.908	0.403	3.089	0.225	−0.714	225.284	1.508	35.2	1.050	1.262
23	0.860	−23498.638	7.374	−5.703	5.961	0.257	1.288	0.269	−0.714	255.310	1.503	34	1.069	0.637
24	0.830	−21452.500	7.183	−5.443	5.831	0.389	2.994	0.264	−0.714	239.311	1.506	35.1	1.036	−1.842
25	0.840	−29171.025	7.129	−5.499	5.468	−0.030	3.920	0.293	−0.712	241.350	1.552	44.2	1.100	1.248
26	0.840	−29171.025	7.102	−5.569	5.499	−0.070	3.714	0.289	−0.708	241.350	1.552	44.2	1.100	−1.003
27	0.810	−28101.400	7.319	−5.536	5.488	−0.049	4.009	0.293	−0.702	227.323	1.559	45	1.120	0.719
28	0.760	−22396.388	7.020	−5.281	5.594	0.313	0.427	0.283	−0.702	239.268	1.540	43.5	1.184	1.294
29	0.660	−29171.025	7.047	−5.407	5.332	−0.075	4.420	−0.226	−0.705	241.350	1.552	44.2	1.100	1.743
30	0.680	−30240.922	7.020	−5.303	5.221	−0.082	4.043	−0.222	−0.713	255.376	1.547	43.6	1.090	2.272
31	0.670	−17266.922	7.047	−5.243	5.765	0.522	2.553	0.378	−0.713	195.258	1.512	34.7	1.023	−1.265
32	0.590	−12104.477	7.809	−5.971	6.340	0.370	1.567	0.179	−0.707	149.233	1.525	35.2	0.938	2.241
33	0.580	−31310.547	7.020	−5.311	5.224	−0.088	4.067	−0.219	−0.712	269.403	1.543	43.1	1.070	2.801
34	0.460	−26669.745	7.238	−5.590	5.669	0.079	3.375	0.267	−0.712	301.380	1.555	39.8	1.093	2.810
35*	0.430	−19280.462	7.211	−5.328	5.872	0.544	1.369	0.265	−0.713	209.242	1.550	45.4	1.175	1.699
36	0.410	−16164.481	7.211	−5.293	5.602	0.309	1.899	0.329	−0.708	179.216	1.564	48	1.163	1.707
37	0.380	−22522.234	7.156	−5.435	5.827	0.391	2.947	0.265	−0.713	253.337	1.503	35	1.023	1.777
38	0.380	−30240.922	7.020	−5.247	5.533	0.286	3.050	0.300	−0.711	255.376	1.547	43.6	1.090	1.777
39*	0.380	−30240.922	7.156	−5.533	5.466	−0.067	3.751	0.285	−0.710	255.376	1.547	43.6	1.090	1.777
40	0.330	−20382.766	7.319	−5.524	5.912	0.388	3.136	0.245	−0.719	225.284	1.507	34.3	1.048	1.042
41	0.230	−21452.609	7.129	−5.410	5.794	0.384	2.979	0.260	−0.713	239.311	1.506	35.1	1.036	1.273
42	0.030	−20382.739	7.238	−5.481	5.872	0.390	3.237	0.249	−0.713	225.284	1.508	35.2	1.050	0.744
43*	0.000	−19312.923	7.319	−5.523	5.914	0.391	3.203	0.253	−0.707	211.258	1.511	35.4	1.067	0.215
44*	−0.030	−19312.787	7.456	−5.719	6.007	0.288	1.449	0.305	−0.710	211.258	1.511	35.4	1.067	−3.080
45*	−0.060	−17266.759	7.292	−5.444	5.849	0.405	3.673	0.308	−0.717	195.258	1.512	34.7	1.023	0.435
46	−0.670	−16196.943	7.292	−5.443	5.848	0.406	3.724	0.307	−0.713	181.232	1.517	36	1.041	0.126

linear regression (MLR) and non-linear regression (MNLR).

The PCA is a mathematical technique used to reduce the dimensionality of a data set consisting of a large number of interrelated variables while retaining as much as possible of the variation present in the data set [18], the method is mostly used as a tool in exploratory data analysis and for making predictive models.

Multiple linear regression (MLR) is a statistical method aimed to establish a mathematical relationship between a property of a given system and a set of descriptors that encode chemical information. Also it serves to select descriptors that are applied as input in multiple non-linear regression (MNLR).

2.4. Validation

The main objective of a QSAR study is to obtain a model with the highest predictive and generalization abilities. In order to evaluate the predictive power of the QSAR models developed, two principals (internal validation and external validation) were performed. For the internal validation the leave-one-out cross-validation (R^2_{CV}) was used to evaluate the stability and the internal capability of the models in the present paper. A high R^2_{CV} value means a high internal predictive power of a QSAR model and a good robustness. Nevertheless, the study of Globarikh [19] indicated that there is no correlation between the value of R^2_{CV} for the training set and predictive ability of the test set, revealing that the R^2cv is still inadequate for a reliable estimate of model's predictive power for all new chemicals. So, the external validation remains the only way to determine both the generalizability and the true predictive power of QSAR models for new chemicals. For this reason, the statistical external validation was applied to the models as described by Globarikh and Tropsha. Roy and Roy [19–21] using a test set.

3. Results and discussion

3.1. Data set for analysis

A QSAR Study was carried out for 46 phenethylamines as reported previously [1,8] in order to establish a quantitative relationship between their structures and the psychotomimetic activity. The values of the 13 calculated descriptors are listed in Table 2.

3.2. Principal component analysis

The principal component analysis (PCA) was performed to the 13 descriptors of the 46 molecules, 13 principal components were obtained (Fig. 2), the first four axis F1, F2, F3 and F4 represent respectively (34.1%; 20.3%; 10.6% and 9.6%) of the total variance and they estimate 74.7% of the total information.

The PCA was performed to identify the correlation between the different descriptors. It is also helpful for understanding the distribution of the compounds [22]. The correlation's matrix of the thirteen descriptors is shown in Table 3.

The correlation coefficients in the obtained matrix provide the information about the high or low interrelationship between the descriptors. Generally good co-linearity ($r > 0.5$) [23] was present between the majority of the variables. A high interrelationship was observed between γ and n ($r = 0.951$). Additionally, to decrease the redundancy existing in our data matrix, the descriptors that are highly correlated ($R \geq 0.8$), were excluded.

3.3. Multiple linear regressions MLR

Based on the twelve remaining descriptors a mathematical linear model was proposed to predict quantitatively the physicochemical effects of substituents on the psychotomimetic activity of the 46 molecules by using backward regression. The best linear model using this method is only one contained five molecular descriptors: the total energy E_T, the energy E_{HOMO}, the energy E_{LUMO}, the dipole moment (DM) and the surface tension (γ).

The following equation represents the QSAR model obtained using the backward regression linear multiple (RLM) method:

$$\log UM = 10.99 - 2.61 \times 10^{-5} \times E_T + 1.29 \times E_{HOMO} - 1.88 \times E_{LUMO} - 0.29 \times DM - 6.60 \times 10^{-2} \times \gamma \quad (1)$$

$$N = 35 \quad R = 0.838 \quad R^2 = 0.712 \quad R^2_{cv} = 0.601$$
$$MSE = 0.143 \quad F = 14.354 \quad P < 0.0001$$

R^2 is the coefficient of determination, F is the Fisher statistic and MSE is the mean squared error. Higher coefficient of determination and lower mean squared error indicate that the model is more reliable. A P smaller than 0.05 means that the obtained equation is statistically significant at the 95% level. The obtained model was cross-validated by its applicable R^2_{cv}value ($R^2_{cv} = 0.601$) using the leave-one-out (LOO) method. A value of R^2_{cv} greater than 0.5 is the basic criteria to qualify a model as valid [19].

The multi-collinearity between the above five descriptors were detected by calculating their variation inflation factors VIF as shown in Table 4; Accordingly, it has been found that the descriptors used in the proposed model have very low-inter-correlation.The VIF [24] was defined as $1/(1 - R^2)$, where R is the coefficient of correlation between one descriptor and all the other descriptors in the proposed model. A VIF value greater than 5.0 indicates that the model is unstable, a

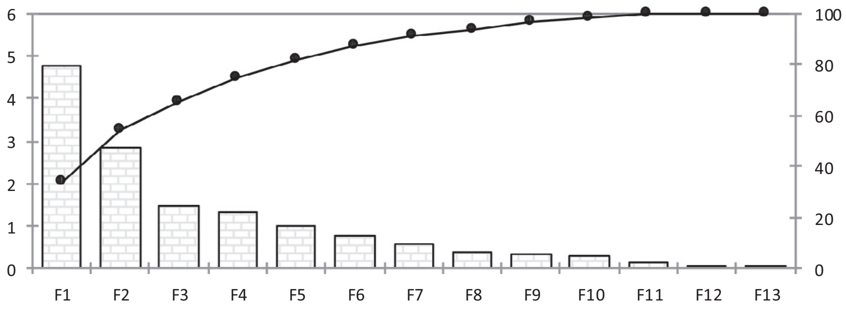

Fig. 2. The principal components and their variances.

Table 3
Correlation matrix between different obtained descriptors.

	log *UM*	E_T	IP	E_{HOMO}	η	E_{LUMO}	DM	Q_p	Q_{min}	MW	*n*	γ	*D*	log *P*
log *UM*	1													
E_T	−0.445	1												
IP	−0.265	−0.084	1											
E_{HOMO}	0.372	0.271	−0.555	1										
η	−0.578	0.259	0.460	−0.633	1									
E_{LUMO}	−0.334	0.607	−0.005	0.253	0.589	1								
DM	−0.259	−0.295	0.057	−0.270	−0.120	−0.432	1							
Q_p	−0.353	0.334	0.022	−0.099	0.606	0.654	−0.308	1						
Q_{min}	0.196	−0.252	0.033	−0.208	−0.147	−0.401	0.026	−0.167	1					
MW	0.342	−0.527	−0.280	0.043	−0.431	−0.493	0.314	−0.326	0.057	1				
n	0.054	−0.351	0.063	−0.150	−0.396	−0.652	0.251	−0.371	0.467	0.198	1			
γ	−0.029	−0.194	−0.026	−0.048	−0.383	−0.530	0.201	−0.310	0.427	0.164	0.951	1		
D	0.200	−0.750	0.098	−0.240	−0.127	−0.409	0.083	−0.111	0.335	0.291	0.627	0.578	1	
log *P*	0.532	−0.202	−0.246	0.396	−0.509	−0.223	−0.112	−0.393	0.052	0.323	0.177	0.116	0.043	1

Table 4
Multicolinearity test.

Variables	E_T	E_{HOMO}	E_{LUMO}	DM	γ
VIF	1.659	1.154	3.159	1.187	2.012

value between 1.0 and 4.0 indicates that the model is acceptable.

Negative values in the regression coefficients show that the indicated variables (E_T, E_{LUMO}, DM and γ) contribute negatively to the value of log *UM*, whereas positive value in the regression coefficient of variable (E_{HOMO}) indicates that the greater the value of the variable, the greater the value of the log *UM*. Put differently, increasing the Total energy E_T, the energy E_{LUMO}, the dipole moment, (DM) and the surface tension (γ) will decrease the log *UM*. While the increase in the energy E_{HOMO} will increase the log *UM* of the phenethylamines.

The correlations of the predicted and observed activities are illustrated in Fig. 3. The descriptors proposed in Eq. (1) by MLR are then used as the input parameters in the multiple nonlinear regressions (MNLR).

3.4. Multiples nonlinear regression (MNLR)

The nonlinear regression model was used also to improve the structure–activity in quantitative manner to evaluate the effect of the substituents on the psychotomimetic activity. Both training set and descriptors selected by MLR were used in this method to build the non-linear model. The best regression performance was selected according to the coefficient of determination R^2 and the mean squared error MSE, a pre-programmed function in the XLSTAT was used to evaluate the nonlinear regression model as follows:

$$Y = a + (bX_1 + cX_2 + dX_3 + eX_4 + \cdots) + (fX_1{}^2 + gX_2{}^2 + hX_3{}^2 + iX_4{}^2 + \cdots).$$

where $X_1, X_2, X_3, X_4, \ldots$ represent the variables, and *a*, *b*, *c*, *d*, ... represent the parameters.

The resulting equation is as follows:

$$\log UM = 88.55 - 8.11\times10^{-5}\times E_{\mathrm{T}} + 22.17\times E_{\mathrm{HOMO}} - 3.69\times E_{\mathrm{LUMO}} - 0.35\times \mathrm{DM} - 1.27\times\gamma - 5.37\times10^{-10}\times E_{\mathrm{T}}^2 + 1.97\times E_{\mathrm{HOMO}}^2 + 6.23\times E_{\mathrm{LUMO}}^2 + 4.87\times10^{-3}\times \mathrm{DM}^2 + 0.015\times\gamma^2 \quad (2)$$

$$N = 35 \quad R = 0.910 \quad R^2 = 0.825 \quad R^2_{\mathrm{cv}} = 0.635 \quad \mathrm{MSE} = 0.105$$

The obtained model Eq. (2) was cross-validated by its applicable $R^2{}_{\mathrm{cv}}$value ($R^2{}_{\mathrm{cv}} = 0.635$) using the leave-one-out (LOO) method. A value of $R^2{}_{\mathrm{cv}}$ greater than 0.5 is the basic criteria to qualify a model as valid [19]. It can be seen clearly from the key statistical indicators, coefficient of determination R^2, mean squared error MSE and, value of $R^2{}_{\mathrm{cv}}$, that the predicting ability of this model is better than that of the linear model (MLR). The enhancement in the predictive ability was due to the involvement of the squared terms in the non-linear model.

Fig. 4 shows the correlation between the predicted and observed log *UM* values.

3.5. Applicability domain

The utility of a QSAR model is its accurate prediction ability for new chemical compounds, so, once the QSAR model is built, its domain of applicability (AD) must be defined. A model is considered valid only within its training domain and only the prediction for new compounds falling within its applicability domain can be considered reliable and not model extrapolations. The most common method to define the AD, it is based on the determination of the leverage value of each compound [21]. The Williams plot (the plot of standardized residuals versus leverage values (h)) is used in the present study to visualize the AD of the QSAR model.

$$h_i = x_i^T (X^T X)^{-1} x_i$$

where the x_i is the descriptor vector of the considered compound, X is the descriptor matrix derived from the training set descriptor values, the threshold is defined as:

$$h* = \frac{3(k+1)}{n}$$

where n is the number of compound in the training set, k is the number of the descriptors in the proposed model, a leverage (h) greater than the threshold (h*) indicates that the predicted response is an extrapolation of the model and, consequently, it can be unreliable.

The Williams plot of the presented MLR model is shown in the Fig. 5, the applicability domain is established inside a squared area within ±2 standard deviation and a leverage threshold h* of 0.51. As shown in the Williams plot the majority of the compounds in the data set are in this area, except one (compound 7) in test set exceeds the threshold and it is considered as an outlier compound, also, compound 1 in the training set is considered as an outlier because it exceeds the crucial hat value. These erroneous predictions could probably be attributed to the presence of the bromine on the para position in chemicals 1 and 7, whereas, the majority of compounds

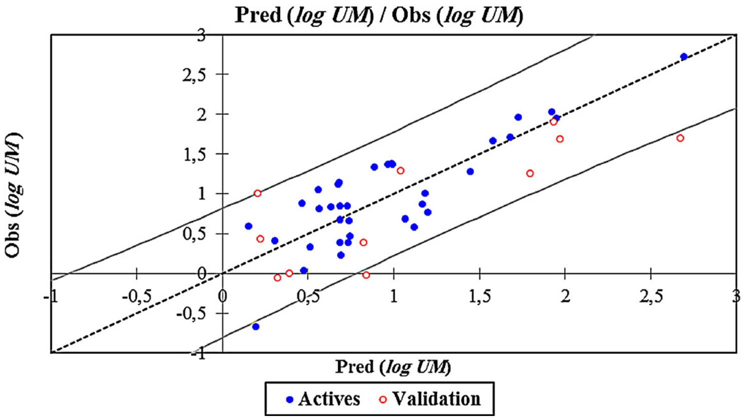

Fig. 3. Graphical representation of calculated and observed activity (log *UM*) values calculated by MLR model.

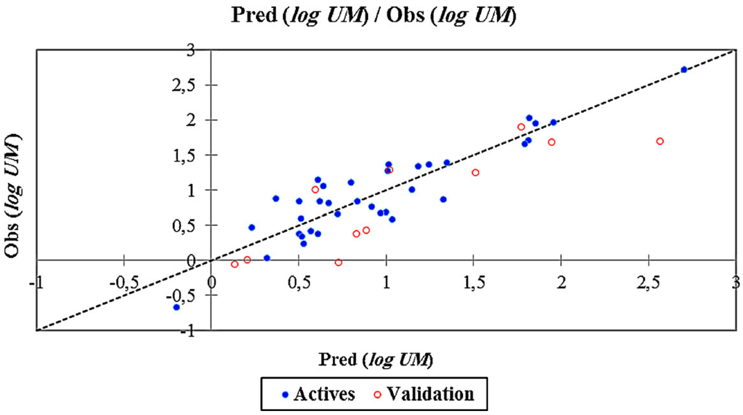

Fig. 4. Graphical representation of calculated and observed activity (log *UM*) values calculated by MNLR model.

are substituted by alkyls, simple ether or thioether groups on this position.

3.6. External validation

To test the prediction ability of the obtained models: MLR and MNLR, it is required the use of a test set for external validation. Thus, the models generated on the training set using 35 phenethylamines were used to predict the activity of the 11 remaining molecules. The parameters of the performance of the generated models are shown in Table 5. It can be seen clearly that the MNLR is statically better than the MLR model.

Among the obtained models, the MNRL model has the highest determination coefficients for the training set ($R^2 = 0.825$) and test set ($R^2_{\text{test}} = 0.746$), also the highest Cross-validation coefficient ($R^2_{\text{CV}} = 0.635$), all that support the applicability of the proposed MNLR prediction model, because the MNLR approach yields better results than those of MLR. However, both the results obtained

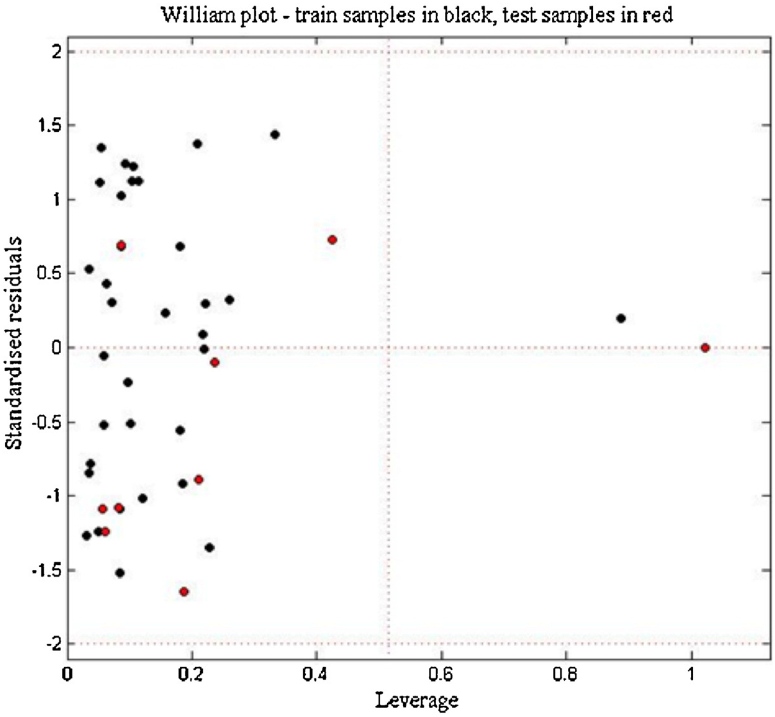

Fig. 5. Williams plot for the training set and external validation for the psychotomimetic activity of phenethylamines compounds, listed in Table 1 ($h^* = 0.51$ and residual limits ±2).

Table 5
the statistical results of MLR and MNLR models with validation techniques.

Method/parameter	R	R^2	R^2_{CV}	R^2_{test}	MSE
MLR	0.838	0.712	0.600	0.635	0.143
MNLR	0.910	0.825	0.635	0.746	0.105

Table 6
Observed values and calculated values of log UM according to different methods.

No.	log *UM* (obs)	log *UM* (pred)	
		MLR	MNLR
1	2.72	2.70	2.69
2	1.96	1.72	1.73
3	2.02	1.92	1.92
4	1.95	1.95	1.95
5*	1.90	1.93	1.77
6	1.71	1.67	1.68
7*	1.69	2.67	2.51
8*	1.68	1.97	1.95
9	1.66	1.57	1.58
10	1.36	0.99	0.99
11	1.33	0.89	0.89
12*	1.25	1.80	1.51
13*	1.29	1.04	1.02
14	1.27	1.45	1.44
15	1.14	0.67	0.68
16	1.36	0.96	0.96
17	1.11	0.68	0.67
18*	1.00	0.20	0.59
19	1.05	0.55	0.56
20	1.00	1.19	1.18
21	1.38	0.98	0.98
22	0.87	0.46	0.46
23	0.86	1.15	1.16
24	0.83	0.63	0.63
25	0.84	0.68	0.68
26	0.84	0.72	0.73
27	0.81	0.56	0.56
28	0.76	1.19	1.20
29	0.66	0.74	0.74
30	0.68	1.06	1.06
31	0.67	0.69	0.68
32	0.59	0.16	0.15
33	0.58	1.11	1.12
34	0.46	0.74	0.74
35*	0.43	0.22	0.89
36	0.41	0.31	0.30
37	0.38	0.68	0.68
38	0.38	0.72	0.73
39*	0.38	0.82	0.83
40	0.33	0.51	0.51
41	0.23	0.69	0.69
42	0.03	0.47	0.47
43*	0.00	0.39	0.21
44*	−0.03	0.84	0.72
45*	−0.06	0.32	0.13
46	−0.67	0.20	0.19

by the MLR and MNLR should be regarded as satisfactory for predicting the psychotomimetic activity using the proposed descriptors.

4. Conclusion

To predict the psychotomimetic activity of substituted phenethylamines compounds, two unambiguous models were developed in this study. Good stability and great prediction ability were achieved by each model. Furthermore, the MNLR results are better, compared to those obtained from the MLR models. So, the MNLR model is considered as an effective tool to predict psychotomimetic activity of substituted phenethylamines based on the proposed descriptors.

The accuracy and predictability of the proposed models were checked based on the domain of applicability and by comparing key statistical indicators, such as the R or R^2 of the obtained models using different statistical tools, as shown in Table 5. To validate these results, a test set was used, as shown in Table 6.

Finally, we concluded that the electronic and physic-chemical descriptors used are able to encode the structural features of the studied compounds, and they could be used successfully with other descriptors for the development of unambiguous predictive QSAR models.

Acknowledgment

We are grateful to the “Association Marocaine des Chimistes Théoriciens” (AMCT) for its pertinent help concerning the programs.

References

[1] L.E. Hollister, Chemical Psychoses: LSD and Related Drugs, Charles C. Thomas, Springfield, IL, 1968.

[2] M. Thakur, A. Thakur, P.V. Khadikar, QSAR studies on psychotomimetic phenylalkylamines, Bioorg. Med. Chem. 12 (2004) 825–831.

[3] A.T. Shulgin, A. Shulgin, PIHKAL. A Chemical Love Story, Transform Press, Berkeley, CA, 1991.

[4] R.A. Glennon, Classical hallucinogens, in: C.R. Schuster, M.J. Kuhar (Eds.), Pharmacological Aspects of Drug Dependence. Handbook of Experimental Pharmacology Series, Berlin, Springer, 1996, pp. 343–371.

[5] R.A. Glennon, Pharmacol. Biochem. Behav. 64 (2) (1999) 251.

[6] R.A. Glennon, in: G.C. Lin, R.A. Glennon (Eds.), Hallucinogens an Update, National Institute on Drug Abuse, Washington, DC, 1994, pp. 4–32.

[7] R.A. Glennon, in: T.L. Lemke, D.A. Williams, V.F Roche, S.W. Zito (Eds.), Hallucinogens, Stimulants, and Related Drugs of Abuse, Foye’s principals of Medicinal Chemistry, Philadelphia, 2008, p. 631.

[8] M. Mracec, L. Kurunczi, T. Nusser, Z. Simon, G. Nàray-Szabo, QSAR study with steric (MTD), electronic and hydrophobicity parameters on psychotomimetic phenylalkylamines, J. Mol Struct. (Theochem) 367 (1996) 139–149.

[9] E. Zvinavashe, T. Du, T. Griff, H.H. van den Berg, A.E. Soffers, J. Vervoort, A.J. Murk, I.M. Rietjens, Quantitative structure–activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* and *Cyprinus carpio*, Chemosphere 75 (2009) 1531–1538.

[10] E. Eroglu, H. Türkmen, A DFT-based quantum theoretic QSAR study of aromatic and heterocyclic sulfonamides as carbonic anhydrase inhibitors against isozyme CA-II, J. Mol. Graph. Model. 26 (2007) 701–708.

[11] S. Arulmozhiraja, M. Morita, Structure–activity relationships for the toxicity of polychlorinated dibenzofurans: approach through density functional theory based descriptors, Chem. Res. Toxicol. 17 (2004) 348–356.

[12] F.A. Pasha, H.K. Srivastava, P.P. Singh, Comparative QSAR study of phenol derivatives with the help of density functional theory, Bioorg. Med. Chem. 13 (2005) 6823–6829.

[13] H. Chermette, Chemical reactivity indexes in density functional theory, J. Comp. Chem. 20 (1999) 129–154.

[14] M.J. Frisch, Gaussian 03 Revision B.01, Gaussian, Inc., Pittsburgh, PA, 2003.

[15] Advanced Chemistry Development, Inc., Toronto, Canada, 2009 www.acdlabs.com/resources/freeware/chemsketch/.

[16] M. Larif, A. Adad, R. Hmamouchi, A.I. Taghki, A. Soulaymani, A. Elmidaoui, M. Bouachrine, T. Lakhlifi, Biological activities of triazine derivatives combining DFT and QSAR results, Arab. J. Chem. (2016), http://dx.doi.org/10.1016/j.arabjc.2012.12.033 465 (in press).

[17] XLSTAT Software, XLSTAT Company, 2013 http://www.xlstat.com.

[18] I.T. Jolliffe, Principal Component Analysis, second ed., Springer, Aberdeen, 2002.

[19] A. Globarikh, A. Tropsha, Beware of q^2!, J. Mol. Graph. Model. 20 (2002) 269–276.

[20] P.P. Roy, K. Roy, On some aspects of variable selection for partial least squares regression models, QSAR Comb. Sci. 27 (2008) 302–313.

[21] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701.

[22] S. Chtita, M. Larif, M. Ghamali, M. Bouachrine, T. Lakhlifi, DFT-based QSAR studies of dibenzo[*a*,*d*]cycloalkenimine derivatives for non competitive antagonists of *N*-methyl-d-aspartate based on density functional theory with electronic and topological descriptors, J. Taibah Univ. Sci. 9 (2) (2014) 143–154.

[23] M. Ghamali, S. Chtita, R. Hmamouchi, A. Adad, M. Bouachrine, T. Lakhlifi, The inhibitory activity of aldose reductase of flavonoid compounds: combining DFT and QSAR calculations, J. Taibah Univ. Sci. 10 (2016) 534–542.

[24] R.M. O’Brien, A caution regarding rules of thumb for variance inflation factors, Q. Quantity 41 (2007) 673–690.